2. (50 points)

3-5 = |-2|
10-2 = |12|
8-6 = |2|
-4-1 = |-5|

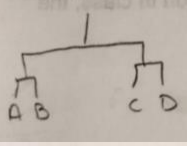
1a. (5 points) X and Y are a pair of 4-dimensional points, where X = (3, 10, 8, -4) and Y = (5, -2, 6, 1). Calculate the **Euclidean distance and the Manhattan distance** between X and Y. You don't need to write a program to do this, but please show your work.

Euclidean: $\sqrt{2^2+12^2+2^2+(-5)^2} = 13.3$ Manhattan: $|2|+|12|+|2|+|-5| = 21$

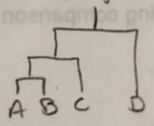
1b. (10 points) Suppose that we have four observations, for which we compute a dissimilarity matrix, given below

	A	B	C	D
Gene A		0.3	0.4	0.7
Gene B	0.3		0.5	0.8
Gene C	0.4	0.5		0.45
Gene D	0.7	0.8	0.45	

Complete



Single

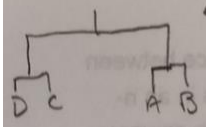


For instance, the dissimilarity between Gene 1 and Gene 2 is 0.3, and the dissimilarity between Gene 2 and Gene 4 is 0.8, etc. On the basis of this dissimilarity matrix, sketch the dendrogram tree that results from hierarchically clustering these four observations using **complete linkage**. Don't worry about the heights of the branches, just provide the topology.

1c. (10 points) Repeat (1b), this time using **single linkage** clustering.

1d. (10 points) Suppose that we cut the dendogram obtained in (1b) such that two clusters result. Which observations are in each cluster? 1) A, B
2) C, D

1e. (10 points) Suppose that we cut the dendogram obtained in (1c) such that two clusters result. Which observations are in each cluster? 1) A, B, C
2) D



1f. (5 points) When we learned about phylogenetic trees, we talked about how the position of the two leaves of the tree being fused can be swapped without changing the meaning of the overall tree. The same is true for hierarchical clustering dendrograms. Draw a dendrogram that is equivalent to the dendrogram in (1b), for which two or more of the leaves are repositioned, but for which the meaning of the dendrogram is the same.